PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **H06B 1/66, H03M 7/30** **A1**	**(11) International Publication Number:** **WO 95/18523** **(43) International Publication Date:** 6 July 1995 (06.07.95)

(21) International Application Number: PCT/IB94/00418

(22) International Filing Date: 13 December 1994 (13.12.94)

(30) Priority Data:
93203652.8   23 December 1993 (23.12.93)   EP
(34) Countries for which the regional or
    international application was filed: AT et al.

(71) Applicant: PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(71) Applicant (for SE only): PHILIPS NORDEN AB [SE/SE]; Kottbygatan 7, Kista, S-164 85 Stockholm (SE).

(72) Inventors: OOMEN, Arnoldus, Werner, Johannes; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL). GROENEWEGEN, Marc, Edwin; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL). VAN DER WAAL, Robbert, Gerbrand; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL). VELDHUIS, Raymond, Nicolaas, Johan; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(74) Agent: STRIJLAND, Wilfred; Internationaal Octrooibureau B.V., P.O. Box 220, NL-5600 AE Eindhoven (NL).

(81) Designated States: JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD AND APPARATUS FOR ENCODING MULTIBIT CODED DIGITAL SOUND THROUGH SUBTRACTING ADAPTIVE DITHER, INSERTING BURIED CHANNEL BITS AND FILTERING, AND ENCODING AND DECODING APPARATUS FOR USE WITH THIS METHOD

(57) Abstract

For encoding sound received as a stream of multibit input samples, from a finite length sequence of input samples, an instantaneous audibility threshold characteristic is determined. Next, a shaped, dither determined, signal is subtracted from the input samples to produce processing samples. Subtracting a dither signal dynamically ignores processing sample bits below the threshold. Next, quantizing by a variable number of bits below the threshold is done, while retaining all sample bits above the threshold. The ignored bits are replaced by the dither signal as buried channel bits at an adjustable number of bits per sample. Therefor, upgraded outputted samples have non-ignored bits and buried channel bits. The noise is difference through shape-filtering a difference between the upgraded samples and the processing samples, which shape-filtering may amend a received white-noise-like signal towards an actual threshold-versus-frequency characteristic.

Method and apparatus for encoding multibit coded digital sound through subtracting adaptive dither, inserting buried channel bits and filtering, and encoding and decoding apparatus for use with this method.

FIELD OF THE INVENTION

The invention relates to a method for encoding sound received as a stream of multibit samples while inserting buried channel bits. Such a method has been described in a publication by M.A. Gerzon and P.G. Craven, 'A High-Rate Buried Channel for Audio CD', preprint 3551 of the Audio Eng. Soc. Conv. Berlin, March 1993. See also International

5 Patent application WO 94/03988, priority August 5, 1992, published February 17, 1994 to the same authors. The first reference bases on a relatively simple way of adding a buried channel through subtractively dithered noise-shaped quantization. Although the added feature allows for enhancing the transmitted high quality sound by a buried channel, the present inventors have discovered that the reference does not fully exploit the potentials of the

10 'subtraction' principle, which omission leads to either a lower than optimum transfer capacity of the buried channel, or alternatively, to a lower than intended residual perceptive quality of the originally high-quality sound, for example CD sound, but not being limited thereto.

SUMMARY TO THE INVENTION

15 In consequence, amongst other things, it is an object of the present invention to improve the characteristics of the buried channel inclusion for optimizing both the residual perceptive quality and the transfer capacity of the buried sound channel. Now, according to one of its aspects, the invention is characterized in that the steps of:
- constituting a finite length sequence of said input samples and in said sequence determining

20 an instantaneous non-uniform audibility threshold-versus-frequency characteristic;
- subtracting a shaped noise signal from said input samples to produce processing samples;
- dynamically ignoring processing sample bits below a bit level associated to said characteristic, through subtracting a dither signal (v) and subsequent quantizing by a variable number of b bits below said bit level, but retaining at least all processing sample bits above

25 said bit level;
- replacing such ignored processing sample bits by said dither signal as buried channel bits featuring an adjustable number of b bits per said processing sample;
- outputting upgraded samples (y) comprising non-ignored processing sample bits and buried

channel bits;

- while producing said noise signal through shape-filtering by a noise shaping filter on a dither-determined difference between said upgraded samples and said processing samples, which shape-filtering amends a received difference signal towards said instantaneous

5   threshold-versus-frequency characteristic.

In particular, the shape-filtering of the difference towards the instantaneous threshold-versus-frequency characteristic allows for creating "spectrum space" at those frequencies where the human hearing system is relatively insensitive. The result is that for a rather sizable length of samples a uniform number of buried channel bits per

10   sample can be injected. For a subsequent sequence, the number of buried channel bits has to be determined again. An extended feature is that next to shaping the characteristic of the main channel, also the characteristic of the buried channel can be shaped in similar manner to optimize transfer capacity. In particular, the dither signal would have to be shaped. This adds a certain amount of complexity to the system.

15   Advantageously, the noise-shaping filter receives a difference signal that approaches a white-noise-like characteristic. In many cases, the requirements to the difference signal are not high with respect to its possible influence on the quality of the main channel, and therefore, the dither may have an arbitrary content. In other situations, the difference may not be correlated with the information of the main channel, or may not even

20   be self-correlated. In the latter situations, the dither is preprocessed to get the appropriate absence of correlation. Such randomizing measures by itself are well-known.

Advantageously, said producing is done by a quantizer filter with a variable characteristic. In this way, improved adaptation to said instantaneous audibility and improved subjective audio quality are reached.

25   Advantageously, said shape-filtering is done with a filter having an overall filter curve compounded from a series of elementary filter curves each positioned at a respective unique grid frequency $\theta_k{}^c$ and having a width Δ_k approximating a local power spectral density of the overall spectrum. This is an extremely straightforward method of modelling the power spectrum density, inter alia allowing said grid frequencies to have non-

30   uniform spacing. In general, this improves the accuracy versus the number of grid frequencies used, and thus speeds up the calculation.

The invention also relates to an encoding apparatus for realizing the encoding and to a decoding apparatus for decoding a signal as acquired though effecting of

the method. Various further advantages are recited in dependent claims.

BRIEF DESCRIPTION OF THE DRAWING

5   These and other aspects and advantages of the invention will be described
more in detail hereinafter with reference to preferred embodiments, and in particular with
reference to the appended drawings that show:

Figure 1 an overall block diagram of a device according to the invention;

Figure 2 a subtractively dithered quantizer for use as a basis for a buried
channel encoder according to the reference enhanced according to the invention;

10   Figure 3 a frequency dependent masked threshold through an exemplary
sound spectrum;

Figure 4 a first exemplary buried channel formation;

Figure 5 a second exemplary buried channel formation;

Figure 6 a simplified CELP encoder without pitch prediction;

15   Figure 7 a noise shaping quantizer;

Figure 8 an elementary filter curve.

BRIEF DISCUSSION OF THE PRINCIPLES

The buried channel technology exploits the fact that an audio or sound
20   signal is often represented by an accuracy expressed by the length of its sample bit string that
is actually too high in terms of perceived audio quality. Therefore, the amount of information
may be reduced for thereby freeing transfer capacity for an additional information service.
The additional information is inserted in the least significant portion of the main signal. To a
conventional receiver this modification of the main signal is irrelevant inasmuch as a human
25   listener will not notice the difference. An enhanced receiver system however, will retrieve
the additional information and produce this on a separate output. According to the invention,
the difference spectrum is shape-filtered for amending a received signal towards an actual
threshold-versus-frequency characteristic. This allows for creating "spectrum space" at those
frequencies where the human hearing system is relatively insensitive. If the difference is
30   white-noise-like, the listener will be even less sensitive to the added channel.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Figure 1 shows an overall block diagram of a device according to the
invention. Block 20 is a source of digital sound that is encoded as a stream of samples that

may consist, for example of 16 bits each, recurring at some 44 kHz. The sound has a predefined bandwidth and can have an arbitrary content such as music, speech, or other. Block 22 constitutes a finite length of these samples, such that those take up a certain interval of time, say, 1024 samples = 0.02 seconds, and therein determines an audibility threshold
5   versus frequency characteristic. The audibility may be determined on the basis of a limited number of key characteristics of the incoming audio. It depends on the instantaneous frequency, on masking effects on one frequency band through influence from another band, it depends on the general or local loudness of the sound, and it may vary between listeners, the last variation being generally ignored, however. The threshold may be determined in ways
10   that by itself are known, and the result will be shown hereinafter. Furthermore, after determination of the threshold, the samples are quantized by ignoring a number of **b** low-significance bits thereof. Block 24 is a source for buried channel bits. The nature of the buried channel may be arbitrary, such as additional comment to the main channel such as displayable subtitles or text, an additional sound channel in a multi-channel sound
15   reproduction, of similar or respective different quality levels, multi-lingual speech service, Karaoke or even video. Also, non-related services are conceivable. However, a particular advantageous usage is to define the buried channel as a MPEG audio channel. By itself this standard has proven useful to provide high-quality audio transfer at an moderate bit-rate. Furthermore, the buried channel itself may consist of two or more sub-channels that are
20   functionally unrelated, although together making up for the buried part of the upgraded samples. In block 26 the ignored sample bits from source 20, or a fraction thereof, starting at the lower significance levels, are replaced by bits from source 24. Moreover, in mutually spaced locations of the stream of upgraded samples, indications are inserted in the buried channel as to what is or will subsequently be the number of ignored bits per sample, and if
25   applicable, when the next indication will occur. For efficient operation the spacing between these indications should be set at optimum value. If the spacing is too small, the overhead increases. If the spacing is too large, the number of ignored bits is too low as seen from the individual samples. The channel 28 that may have transmitting, storing or further quality, forwards the upgraded samples to receiver 30. Receiver 30, on the basis of these received
30   indications can separate the standard part of the samples from the buried channel bits. The standard part of the samples is forwarded to decoder 32 that represents the audio in standard manner, wherein the substituted bits are maintained, as representing sub-audible noise. The buried channel bits are forwarded to a subsidiary decoder that has been programmed for correct processing thereof. Another set-up is that the buried channel decoder receives the

whole samples as produced at a digital output of a channel receiver and extracts the buried
channel information therefrom, while ignoring the standard channel. If, on the other hand a
normal non-upgraded receiver has been provided for the channel, this will process the
upgraded samples as if they were normal samples. This eventually ends in an analog audio

5   amplifier that feeds some kind of speaker. The buried channel bits, being generally
uncorrelated to the main channel bits, now directly represent some kind of noise that remains
below the intended auditory threshold.

            Figure 2 shows a subtractively dithered quantizer for use as a basis for a
buried channel encoder according to the reference, as enhanced with certain features

10   according to the invention. On input 52 a b-bit dither signal v is input in a way to be
discussed hereinafter. Elements 54, 58, 60, 62 are digital adders-subtractors of an
appropriate bit width. Element 56 is a quantizer that reduces the accuracy of the exemplary
16-bit received signal x to a lower number of 16-b bits by means of truncation. Such a
feature by itself has been described in S.P. Lipshitz et al, "Quantization and Dither: A

15   theoretical Survey", J. Audio Eng. Soc. Vol. 40, no.5, pp.355-375, May 1992. The usage of
a b-bit dithered signal v, if the lack of correlation with the main channel is sufficient,
ensures that the quantization error e remains spectrally white and statistically independent
from the input signal x, which is preferable for perceptivity reasons. The dithered signal may
be a randomized version of the buried channel signal, without adding or sacrificing

20   information. Such randomization can be reversed without the need for resynchronization. It is
recognized that the random character is specified relative to the main channel, as well as
within the buried channel itself. If within the context of the buried channel itself, the signal
is well-structured, it may be randomized by conventional means. The same dither signal v is
added in element 54 to form the compatible output signal y at output 66, for storage,

25   transmission or further processing. Block 64 is a noise shaping filter and receives the
difference between the compatible output signal y and the input signal before introduction of
the dithered signal v, as produced by subtractor 62. The output signal of noise shaping filter
64 is fed back to subtractor 60 that in its turn receives the original audio signal x. It has been
found that the noise loudness can be decreased by about 16 dB with only a 9-th order FIR

30   (finite impulse response) filter. This approach will make the noise level of a 2-3 bits per
sample buried channel signal of comparable loudness as the inherent noise floor in the CD
signal. The filter characteristic H(Z) should be such that y, defined as

$$y = x + | 1 - H(Z) |^2 . \Delta^2 / 12$$

should be changed with respect to x by an amount that is subjectively inconsequential to the listener, $\Delta = 2^b$ being the elementary step size. Now, the transfer rate of the buried channel depends on the coarseness of the requantization operation. In this respect, p.2, 1.40 ff and p.13, fifth para, of the first Gerzon at al reference indicate that the requantization may be

5   made more coarse when the resulting error is masked by a high level main audio signal. The present inventors, on the other hand, have discovered an even more effective way to increase the transfer rate of the buried channel, namely by using the frequency-dependent sensitivity of the human hearing system. A further refinement to the arrangement of Figure 2 is the buffer 50 that may temporarily store the data supplied by the buried channel 68. In view of

10   the non-uniform rate of the buried channel at output 66, the buffer may have some kind of feed-back organization that keeps its filling degree more or less constant. If the buffer gets too empty, the full capacity of the buried channel may be surrendered in part. If the buffer gets too full, there are various strategies: one is lowering the feed rate from source 68. A more drastic one if more than one buried sub-channel is present, is the surrendering of the

15   least important sub-channel therefrom. If the subchannel represents moving video, it could temporarily be reduced to a sequence of stills. Various modifications of the arrangement of Figure 2 are self-evident: for example, certain of the addition devices may be changed to subtraction devices. This would slightly complicate the hardware, because of the necessary propagating of borrows. Especially with certain sample notation systems, the impact is

20   minimal, however.

In this respect, Figure 3 shows a frequency dependent masking threshold through a first exemplary sound spectrum. Figure 4 shows a first exemplary buried channel formation, based on zero sound input.

In Figure 4, for a standardized or presumed human hearing system, curve

25   4 gives the audibility level, as based on single frequencies. For calculatory reasons, curve 4 has been simulated (note the straight parts thereof), but it closely follows the natural phenomena. At approximately 4000 Hz this threshold is lowest, whilst being much higher at either higher or lower frequencies. Now, trace 1 indicates the flat CD noise floor, that is given as 10^{10} log (1/12 x 22050) dB. Now, although curve 4 gives the audibility threshold

30   for single frequencies, for noise the audibility effects are much higher, and its frequency characteristic should lie much lower than curve 4. Now, curve 2 is the spectrum of the flat noise of curve 1, shaped to get approximately the same frequency dependence as the threshold of curve 4. It has been found experimentally that neither the few dB deviations from the exact approximation, nor the markedly flatter shape above some 15 kHz have a

negative influence on the overall performance. Curve 3 is identical to curve 2, but relatively shifted upward over a distance of b*6 dB, wherein in Figure 4, b=2. This implies a buried channel of two (2) bits wide per sample. It has been found that the distance between curves 3 and 4 keeps the added information unheard. The minimum distance between curves 1 and 4

5   is $10^{10} \log(660)$, which corresponds to the critical bandwidth around 4kHz. The design is preferably made with the help of an auxiliary curve that indicates the integrated power of the spectrum of curve 3, and which may not touch curve 4; for simplicity, this auxiliary curve has not been shown.

Figure 3 shows the influence of an actual sound spectrum on the shape of

10   the audibility threshold. Now curve A is the actual sound spectrum. Curve B again shown the audibility threshold for single tones. Clearly, the valley-like character of curve 4 in Figure 4 has disappeared.

Figure 5 shows a second exemplary buried channel formation. Here the simulated audio spectrum, that may however have a different shape from curve A in Figure

15   3, cause the masked audio threshold to lie at an approximate level of 40 dB with a shallow dip of some 10 dB at 13 kHz. For clarity, the spectrum of the audio itself has not been shown. As should be clear from figure 3, each separate spectrum peak may raise the masked threshold over a frequency width in the order of the critical frequency, which generally causes the smooth appearance of the threshold curve. Again, the influence of the high-

20   frequency range over 16 kHz has been ignored. Now, curve 1 is the same as in Figure 4. Curve 2 is again the spectrum of the flat noise, shaped to get approximately the same appearance as the masked threshold of curve 4; in this case, the overall shape of curve 2 is much flatter than in the preceding Figure. Curve 3 is identical to curve 2, but relatively shifted upward over a distance of b*6 dB, wherein in Figure 5, b=6. This implies a buried

25   channel of six bits wide per sample. It has thus been shown that for higher sound energy the width of the buried channel may increase. It has further been shown that the shape of the sound spectrum is crucial in many cases. With the shape of curve 2 in Figure 4 applied in Figure 5, the improvement would have been much less in the latter Figure. The approach of the preceding Figures has been based on the full frequency spectrum. In certain audio

30   systems, the spectrum has been distributed in subbands, wherein the respective subbands carry largely independent audio signals. In such situation, the method and apparatus of the present invention can be applied separately for any subband or subband combination separately. At the price of a somewhat higher complexity, this would further raise the transfer data rate of the buried channel.

COMPUTATION OF NOISE WEIGHTING FILTERS AND NOISE SHAPING FILTERS
FROM MASKED TARGET LEVELS

Hereinafter, a preferred embodiment for calculating a compound filter
curve for filter 64 in figure 2 is presented. A relevant publication in this field is E.
5   Ordentlich and Y. Shoham, Low-delay code-excited linear-predictive coding of wideband
speech at 32 kbps, Proc. ICASSP-91, pp. 9-12, Toronto, 1991. By itself, the technique
presented infra is suitable for various applications in different fields, such as MPE, RPE, and
CELP. Therein, an excitation sequence (MPE, RPE) or excitation vector (CELP = codebook
excited linear prediction) is selected on the basis of a weighted mean square error criterion.
10   In such a coder, short output sequences are generated from a number of
excitation sequences or vectors. The generated output sequences are compared with the
original input sequences. The criterion for comparison is the weighted mean-squared error.
This means that the difference between input and generated output is passed through a noise
weighting filter. The power of the filtered difference sequence is then estimated. This power
15   is called the weighted mean-squared error. The excitation sequence yielding the minimum
weighted mean-squared error is selected.

Figure 6 shows how a weighting filter is used in a CELP coder. From the
incoming speech signal x[i] the box LPC analysis computes the prediction coefficients a_1, ...,
a_p, the coefficients for the weighting filter and a gain factor. The codebook contains a
20   number of excitation vectors. The vector length is N. During selection all vectors are
multiplied with the gain factor and passed through an analysis filter. This results in a
sequence of N samples denoted by $\hat{x}[i]$. An error sequence is formed by subtracting N input
samples x[i] from N samples $\hat{x}[i]$. The error sequence is passed through the weighting filter.
The weighted mean-squared error, which is the short-term power of the weighted error
25   sequence, is computed. The selection box selects the code vector that results in the lowest
weighted mean-squared error. Gain factor, prediction coefficients and the index of the
vectors are transmitted to the decoder.

In this context, an excitation vector is considered as an excitation
sequence, therefore only the expression excitation sequence will be used.
30   The commonly used weighting filters are based on the prediction
coefficients (LPC coefficients) a_1, ..., a_p of the speech signal. A possible form of this filter
is

$$W(z) = \frac{1 + \sum_{k=1}^{p} a_k \gamma_1^k z^{-k}}{1 + \sum_{k=1}^{p} a_k \gamma_2^k z^{-k}} \cdot \frac{1}{1 + \sum_{k=1}^{2} p_k \delta^k z^{-k}} \tag{1}$$

The coefficients p_1 and p_2 are found by applying LPC analysis to the first three autocorrelation lags of the sequence. The coefficients δ, γ_1 and γ_2 control the amount of weighting at the position of the formants. They are tuned such that good perceptual

5   performance is obtained. Advantageous values are:

$$\delta = 0.7, \gamma_1 = 0.95, \gamma_2 = 0.8.$$

Other similar forms are useful as well. For a well-chosen codebook, the power spectral density function of the coding error is proportional to

$$\left| \frac{1}{W(\exp(j\theta))} \right|^2. \tag{2}$$

10

NOISE SHAPERS

         The function of a noise shaper is to give spectral shaping to quantization noise. Figure 7 shows the basic diagram of a noise shaper. It can be shown that after decoding the power spectral density function of the quantization noise is given by

15   $$S(\exp(j\theta)) = |1 + F(\exp(j\theta))|^2 \frac{1}{12}\Delta^2, \tag{3}$$

where Δ is again the quantization step size. A commonly used filter $F(z)$ in linear predictive coding of speech is

$$F(z) = \sum_{k=1}^{p} a_k \gamma^k z^{-k}, \tag{4}$$

with $\gamma < 1$. In that case the power spectral density function of the quantization noise is

20   given by

$$S(\exp(j\theta)) = \frac{|1 + F(\exp(j\theta))|^2}{|1 + \sum_{k=1}^{p} a_k \exp(-j\theta)|^2} \frac{1}{12}\Delta^2 = \frac{|1 + \sum_{k=1}^{p} a_k \gamma^k \exp(-j\theta)|^2}{|1 + \sum_{k=1}^{p} a_k \exp(-j\theta)|^2} \frac{1}{12}\Delta^2, \tag{5}$$

The aim of a weighting filter and of a noise shaper is spectral shaping of the coding error in such a way that distortion is perceptually least disturbing. In music coding several methods estimate the masking level of quantization noise in frequency bands. This level is called

masked target level. The goal of these methods is also to obtain a distortion that is perceptually least disturbing. However, they have a better psychoacoustical foundation than the weighting filter of (1) or the noise shaper of (4) and will therefore result in a better approximation of the power spectral density function of masked noise.

5   The following combines the weighting filter technique or the noise shaping technique with the computation of masked target level. The weighting filters or noise shapers that are thus obtained are better than the known ones because their transfer functions correspond better to the spectral shape of the masked noise.

10   TARGET LEVELS AND FREQUENCY BANDS

          It is assumed that a set of target levels t_1, ..., t_m is computed in advance, for instance, through measurements discussed with respect to Figures 3, 4, 5. The target levels represent noise powers in frequency bands at masking threshold. These frequency bands must be adjacent and cover the range from zero to half the sampling frequency.

15   Normalized frequencies θ are used, therefore

$$-\pi \leq \theta \leq \pi.$$

The corresponding audio normalized frequency f follows from

$$f = \frac{\theta}{2\pi}f_s,$$

where f_s is the sampling rate.

20   In the following the bandwidths may be chosen arbitrarily. In practice, critical bands or equally spaced bands will be used. The lower edge, upper edge and centre frequency of the frequency band corresponding to masked target level t_k are denoted by

θ_k^l, θ_k^u and θ_k^c, respectively.

25   RECONSTRUCTION OF POWER SPECTRAL DENSITY

          A smooth power spectral density function (psd): $S(\exp(j\theta))$ can be derived from the masked target levels by associating with the kth frequency band a psd-reconstruction function $S_k(\theta)$. The psd then follows from

$$S(\exp(j\theta)) = \sum_{k=1}^{m} t_k S_k(\theta). \tag{6}$$

There are some constraints for the psd-reconstruction functions. In the first place, the psd must be non-negative. This implies that

$$S_k(\theta) \geq 0, \ k = 1 \dots, m \tag{7}$$

In the second place, if power preservation, i.e.

5
$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S(\exp(j\theta))d\theta = \sum_{k=1}^{m} t_k, \tag{8}$$

is required, then

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S_k(\theta) \ d\theta = 1, \ k = 1, \dots, m. \tag{9}$$

A stepwise approximation of the psd is obtained by choosing psd-reconstruction functions

10
$$S_k(\theta) = \frac{\pi}{\Delta_k}, \ \theta_k^l \leq |\theta| < \theta_k^u, \ 0, \ otherwise, \tag{10}$$

where $\Delta_k = \theta_k^u - \theta_k^l$. Stepwise approximations do not result in a smooth psd. For that a raised-cosine psd-reconstruction function is better suited. This is given by

$$S_k(\theta) = \frac{\pi}{2\Delta_k}(1+\cos(\frac{\pi}{\Delta_k}|\theta|-\theta_k^c)), \ \theta_k^c - \Delta_k \leq |\theta| < \theta_k^c + \Delta_k, \ 0, \ otherwise \tag{11}$$

Figure 8 shows such an elementary filter curve. Also the raised-cosine psd
15   reconstruction function has its limitations. It cannot for instance reconstruct a flat psd if the frequency bands are unevenly spaced. This can be improved by using different upper and lower slopes. The choice of a psd-reconstruction function is determined by the desired spectral behaviour of the masked noise.

20   APPROXIMATION OF POWER SPECTRAL DENSITY

The reconstructed psd $S(\exp(j\theta))$ is approximated by an all-pole spectrum

$$\hat{S}(\exp(j\theta)) = \frac{\sigma^2}{|1 + \sum_{k=1}^{q} b_k\exp(-jk\theta)|^2}, \tag{12}$$

where q is the order of the all-pole spectrum. This results in a weighting filter with transfer function

$$B(z) = 1 + \sum_{k=1}^{1} b_k z^{-k}. \tag{13}$$

The weighting filter is an FIR filter, in contrast to the filter of (1). In the following the b_1, ..., b_q are computed from the t_1, ..., t_m, by minimizing

$$Q(b_1,...,b_q) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\exp(j\theta)) |B(\exp(j\theta))|^2 \, d\theta \tag{14}$$

5   as a function of b_1, ..., b_q. In (14), $S(\exp(j\theta))$ follows from (6). By way of summarizing: it is required to approximate the spectrum $S(\theta)$, wherein θ is a normalized frequency. Now, the function B is the inverse function of F(eq.23) and Q is a constant. Now, computing derivatives

$$\frac{\partial Q(b_1, ..., b_q)}{\partial b_n}, \quad n = 1, ...,q,$$

10   and setting them equal to zero leads to the following set of equations

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=1}^{m} t_k S_k(\theta) \left[1 + \sum_{l=1}^{q} b_l \exp(-jl\theta) \right] \exp(jn\theta) d\theta = 0, \quad n=1,...,q. \tag{15}$$

Or,

$$\sum_{l=1}^{q} b_l \sum_{k=1}^{m} t_k \frac{1}{2\pi} \int_{-\pi}^{\pi} S_k(\theta) \exp(j\theta(n-l)) d\theta =$$

$$-\sum_{k=1}^{m} t_k \frac{1}{2\pi} \int_{-\pi}^{\pi} S_k(\theta) \exp(jn\theta) d\theta, \quad n=1,...,q \tag{16}$$

Define

15   $$g_{k,n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_k(\theta) \exp(jn\theta) d\theta, \quad k=1,...,m, \quad n=1,...,q \tag{17}$$

and

$$p_n = \sum_{k=1}^{m} t_k g_{k,n}. \tag{18}$$

The $g_{k,n}$ can be computed in advance from the psd-reconstruction functions and stored in an m x q matrix. On substitution of these results into (16) one obtains the set of equations

$$\sum_{l=1}^{q} p_{n-l} b_l = -p_n, \quad n = 1,...,q.\tag{19}$$

This is a symmetrical, positive-definite Toeplitz system that is identical to the Yule-Walker equations, known from linear predictive coding. Define the q x q matrix R by

$$r_{i,j} = p_{i-j}, \quad i,j = 1, ..., q,$$

5 and the q vector r by

$$r_i = p_i, \quad i = 1, ..., q.$$

This leads to

$$Rb = -r,\tag{20}$$

where the q vector b contains the coefficients $b_1, ..., b_q$. The set (19) or (20) is easily solved

10 by the known Levinson-Durbin algorithm.

EXAMPLES OF $g_{k,n}$

For the stepwise approximation of $S(\exp(j\theta))$ the $g_{k,n}$ are given by

$$g_{k,n} = \cos(n\theta_k^c) \frac{\sin(n\frac{\Delta_k}{2})}{n\frac{\Delta_k}{2}}.\tag{21}$$

15 For the raised-cosine approximation of $S(\exp(j\theta))$ the $g_{k,n}$ are given by

$$g_{k,n} = \cos(n\theta_k^c) \frac{\sin(n\Delta_k)}{n\Delta_k} \frac{(\frac{\pi}{\Delta_k})^2}{(\frac{\pi}{\Delta_k})^2 - n^2}.\tag{22}$$

The coefficients $b_1, ..., b_q$ can be directly applied in a weighting filter W(z), as shown in Figure 7. In that W(z) = B(z), with B(z) defined in (13). In case of a noise shaper, the F(z) follows from

20 $$1 - F(z) = \frac{1}{B(z)},\tag{23}$$

therefore

$$f(z) = \frac{\sum_{k=1}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{q} b_k z^{-k}}.\tag{24}$$

WO 95/18523   PCT/IB94/00418

14

CLAIMS:

1.   A method for encoding a sound representation received as a stream of multibit input samples, characterized by the steps of:

- constituting a finite length sequence of said input samples and in said sequence determining an instantaneous non-uniform audibility threshold-versus-frequency characteristic;

5 - subtracting a shaped noise signal from said input samples to produce processing samples;

- dynamically ignoring processing sample bits below a bit level associated to said characteristic, through subtracting a dither signal (v) and subsequent quantizing by a variable number of b bits below said bit level, but retaining at least all processing sample bits above said bit level;

10 - replacing such ignored processing sample bits by said dither signal as buried channel bits featuring an adjustable number of b bits per said processing sample;

- outputting upgraded samples (y) comprising non-ignored processing sample bits and buried channel bits;

- while producing said noise signal through shape-filtering by a noise shaping filter on a

15 dither-determined difference between said upgraded samples and said processing samples, which shape-filtering amends a received difference signal towards said instantaneous threshold-versus-frequency characteristic.

2.   A method as claimed in Claim 1, wherein said noise-shaping filter receives a difference signal that approaches a white-noise-like characteristic.

20 3.   A method as claimed in Claim 1 or 2, wherein said producing is done by a noise-shaping filter with a variable characteristic.

4.   A method as claimed in Claims 1, 2 or 3, and featuring detecting a musical transient in a particular time interval, and upon such detecting setting the value of b in that interval to a value at least co-determined from a neighbouring time interval not having

25 that musical transient.

5.   A method as claimed in any of Claims 1 to 4, and featuring temporary buffering of data to be used as buried channel data before said replacing in a buffer and through a time varying rate control time-wise equalizing a filling degree of said buffer.

6.   A method as claimed in any of Claims 1 to 5, wherein said buried channel

data is received as an MPEG audio channel.

7.   A method as claimed in any of Claims 1 to 6, wherein said shape-filtering
is done with a filter having an overall filter curve compounded from a series of elementary
filter curves each positioned at a respective unique grid frequency θ_k^c and having a width Δ_k
approximating a local power spectral density of the overall spectrum.

8.   A method as claimed in Claim 7, wherein said grid frequencies have non-
uniform spacing.

9.   A method as claimed in any of Claims 1 to 8 and applied separately to
respective frequency subbands that are coexistent in a frequency spectrum of said sound
representation.

10.   An encoding apparatus for effecting a method as claimed in any of Claims
1 to 9.

11.   A decoding apparatus for decoding a signal of upgraded samples produced
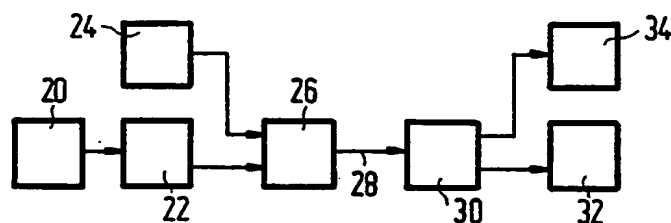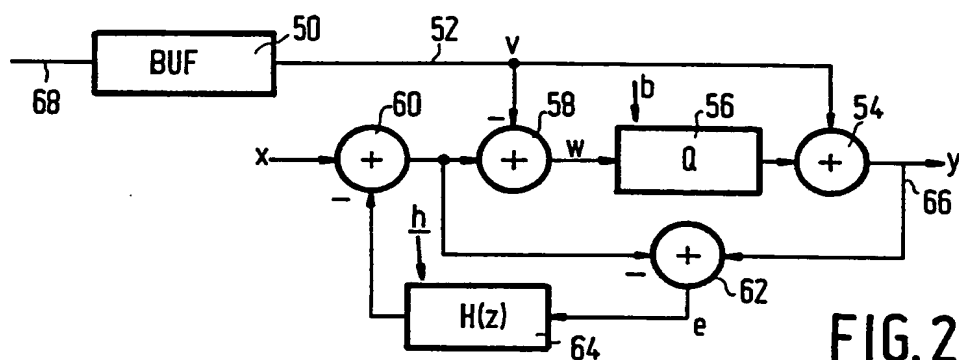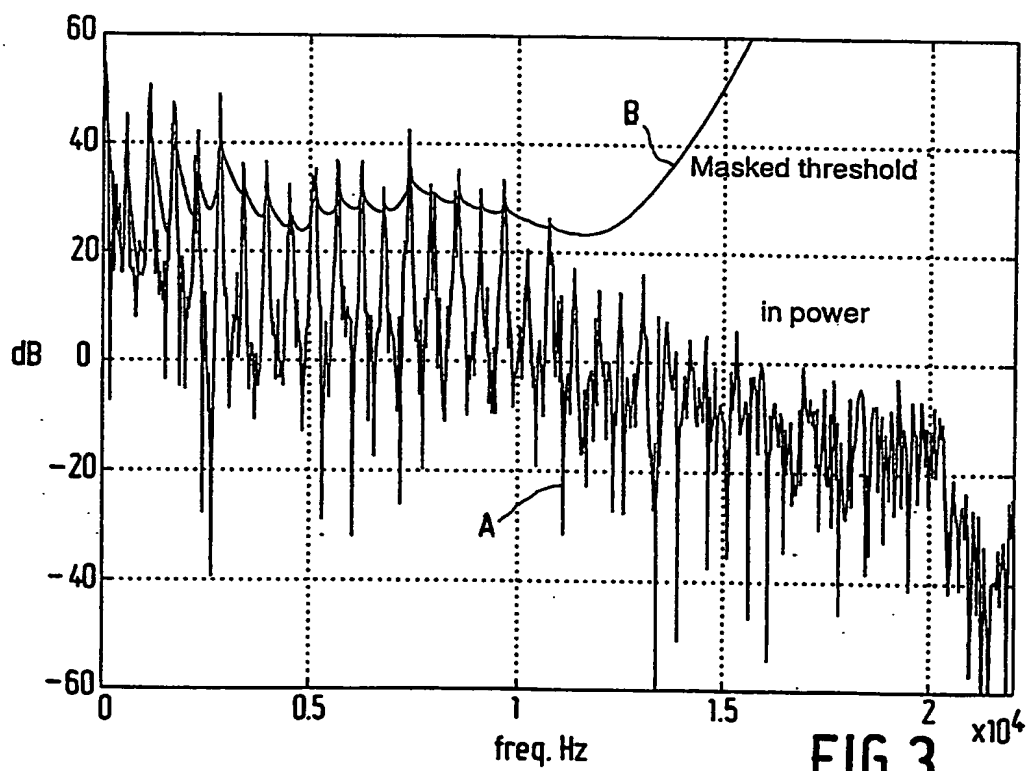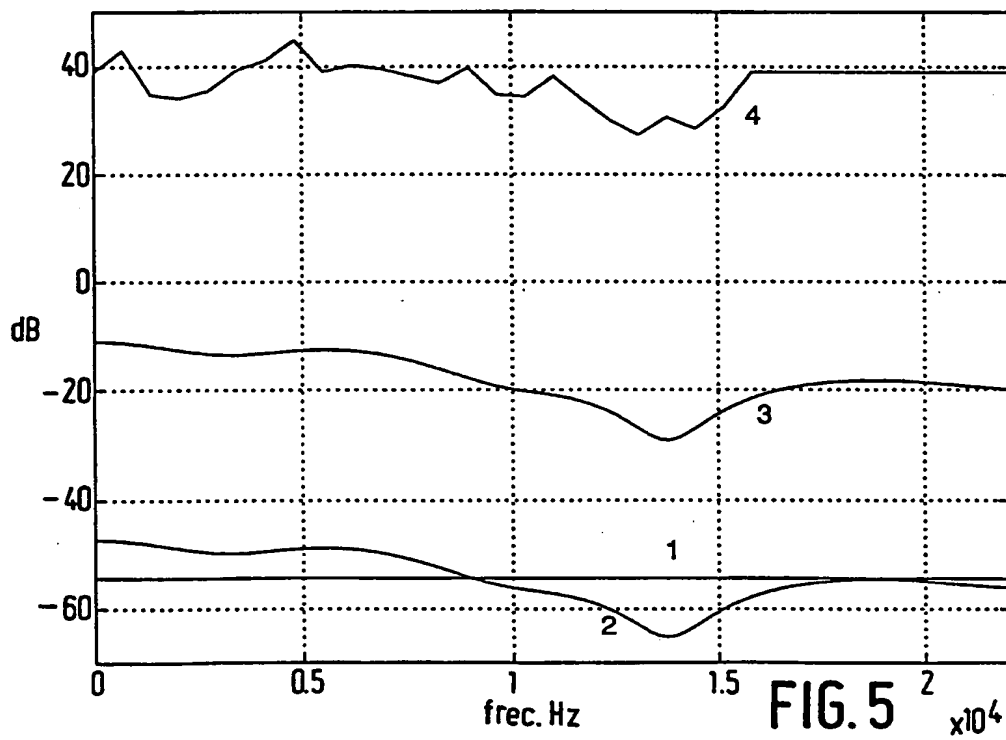through effecting a method as claimed in any of Claims 1 to 9.

1/3

FIG.1

FIG.2

FIG.3

FIG. 4

FIG. 5
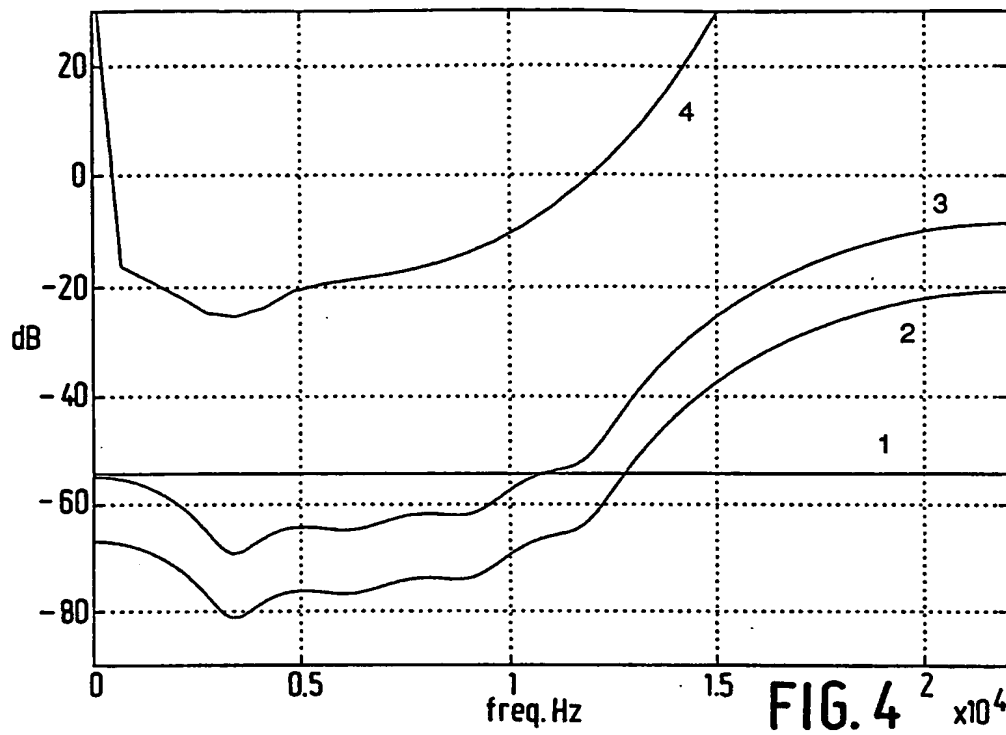
3/3
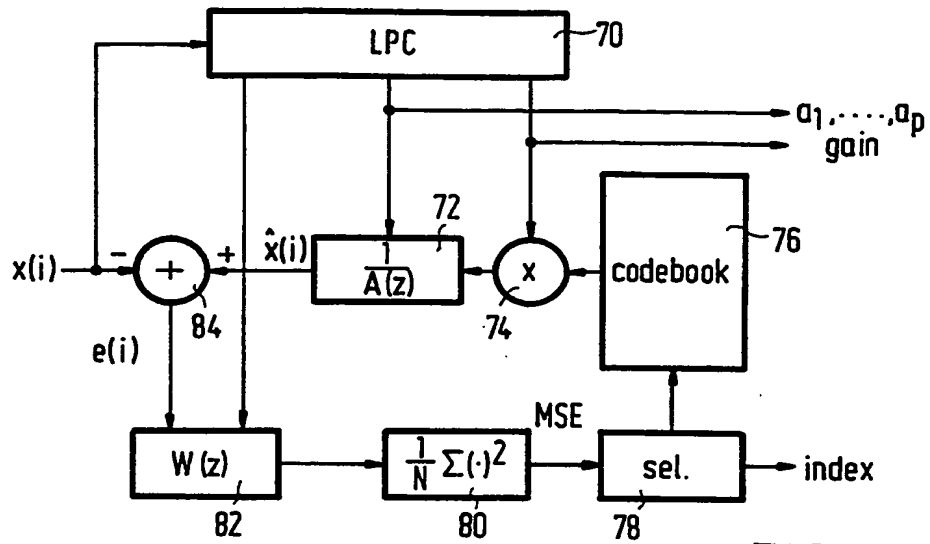
FIG.6
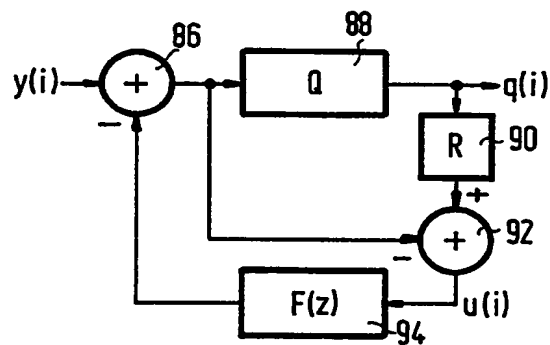
FIG.7
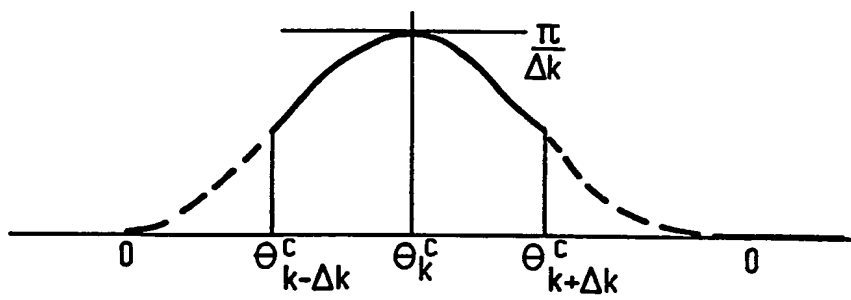
FIG.8

INTERNATIONAL SEARCH REPORT

	International application No.
	PCT/IB 94/00418

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: H04B 1/66, H03M 7/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: H03M, H04B, G11B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	WO, A1, 9418762 (GERZON MICHAEL), 18 August 1994 (18.08.94), page 3, line 34 - page 10, line 9; page 30, line 20 - page 37, line 22 --	1-11
A	US, A, 5185800 (YANNICK MAHIEUX), 9 February 1993 (09.02.93), column 2, line 61 - column 4, line 57 --	1-11
P,A	WO, A2, 9403988 (GERZON MICHAEL ANTHONY), 17 February 1994 (17.02.94), page 3, line 19 - page 6, line 30, cited in the application --	1-11

[X] Further documents are listed in the continuation of Box C.   [X] See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	
"E" erlier document but published on or after the international filing date	"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	
"O" document referring to an oral disclosure, use, exhibition or other means	"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"P" document published prior to the international filing date but later than the priority date claimed	"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 June 1995	1 3 -06- 1995
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. + 46 8 666 02 86	Authorized officer Bo Gustavsson Telephone No. + 46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB 94/00418

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO, A1, 9222060 (PACIFIC MICROSONICS, INC), 10 December 1992 (10.12.92), page 3, line 4 - page 6, line 12; page 13, line 1 - page 15, line 36	1-11
A	EP, A2, 0540330 (SALON TELEVISIOTEHDAS OY), 5 May 1993 (05.05.93), see whole document	1-11

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A1- 9418762	18/08/94	NONE		
US-A- 5185800	09/02/93	DE-D-	69015105	00/00/00
		EP-A,B-	0423050	17/04/91
		FR-A,B-	2653280	19/04/91
		JP-A-	3167927	19/07/91
WO-A2- 9403988	17/02/94	NONE		
WO-A1- 9222060	10/12/92	AU-A-	2161892	08/01/93
		CA-A-	2110182	10/12/92
		EP-A-	0586565	16/03/94
		JP-T-	6509201	13/10/94
EP-A2- 0540330	05/05/93	FI-B-	89439	15/06/93

THIS PAGE BLANK (USPTO)